Занятие 7. Лекция. Статистические тесты для сравнения групп

Обсудим 4 метода

- Mann-Whithney
- Wilcoxon
- Kruskal-Wallis
- Friedman

Тест Уилконсона

Тест для случая парных повторных наблюдений

Пример

Пациенты проходили лечения, призванное уменьшить склонность к суициду. До лечения была склонность X_i, после Y_i

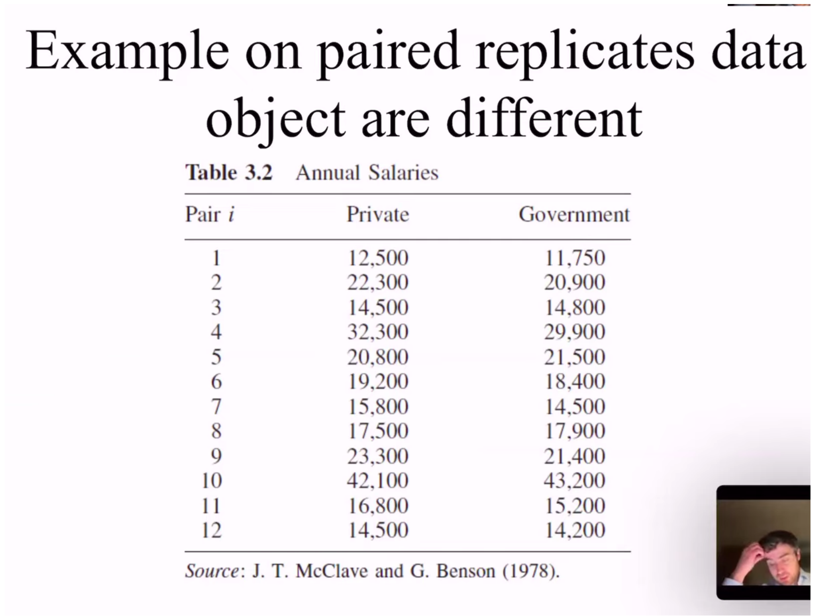
Example on paired replicates objects are the same

Table 3.1 The Hamilton Depression Scale Factor IV Values

Patient i	X_i	Y_i
1	1.83	0.878
2	0.50	0.647
3	1.62	0.598
4	2.48	2.05
5	1.68	1.06
6	1.88	1.29
7	1.55	1.06
8	3.06	3.14
9	1.30	1.29

Source: D. S. Salsburg (1970).

Имеет смысл взять разности. Другой пример: зарплаты в частном и государственном секторе для одних и тех же профессий.



Example on paired replicates data object are different

Table 3.2 Annual Salaries

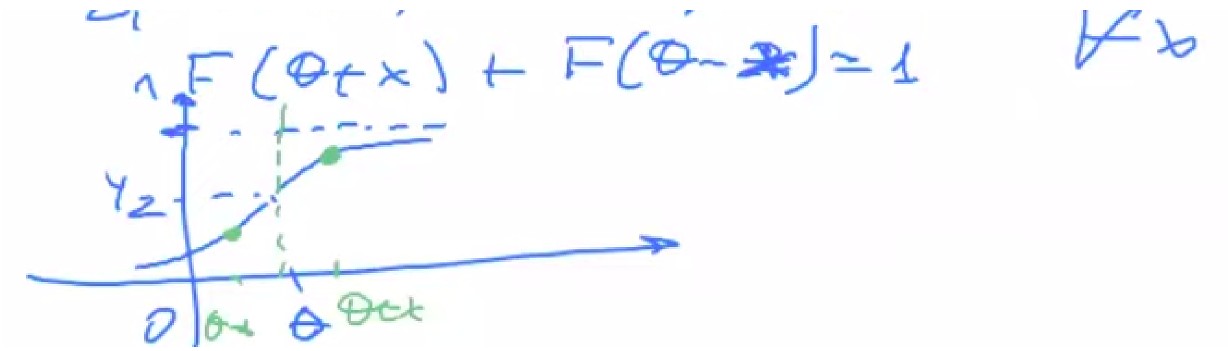
Pair i	Private	Government
1	12,500	11,750
2	22,300	20,900
3	14,500	14,800
4	32,300	29,900
5	20,800	21,500
6	19,200	18,400
7	15,800	14,500
8	17,500	17,900
9	23,300	21,400
10	42,100	43,200
11	16,800	15,200
12	14,500	14,200

Source: J. T. McClave and G. Benson (1978).

Парные наблюдения $\quad\Leftrightarrow\quad$ мы можем что-то померить два раза

Формализация

Есть наблюдения $(X_1, Y_1), \ldots, (X_n, Y_n)$. Выборки зависимы. Их разности: $Z_i = X_i - Y_i$.
Предположение: $Z_i - \text{i.i.d.}$ и симметрично расположены относительно нуля



Гипотеза:

$$
\begin{aligned}
\mathcal{H}_0 &: \quad \theta = 0 \\
\mathcal{H}_1 &: \quad \theta \neq 0 \vee \theta < 0 \vee \theta > 0
\end{aligned}
$$

Идея теста: отсортируем z_i по возрастанию модуля:

$$
|z_{(1)}| \leqslant |z_{(2)}| \leqslant \cdots \leqslant |z_{(n)}|
$$

Вычислим такое число:

$$
W = \sum_i \operatorname{rank} z_i \cdot \mathbb{I}[z_i > 0]
$$

Пример:

X	-2	0	8
Y	3	-4	9
Y−X	5	-4	1
R	3	2	1

$$W = 3 + 1 = 4$$

Все возможные случаи:

точки:	W	H_0
(1 2 3)	6	1/8
(0, 2, 3)	5	1/8
(1, 0, 3)	4	1/8
(1, 2, 0)	3	
(0, 0, 3)	3	.
(0, 2, 0)	2	
(1, 0, 0)	1	.
(0, 0, 0)	0	1/8

Асимптотическая теорема

Для небольших значений n можем явно посчитать. Иначе используем аппроксимацию:

$$H_0 \Rightarrow \frac{W - \frac{n(n+1)}{4}}{\sqrt{\frac{n(n+1)(2n+1)}{24}}} \to N(0,1)$$

Mann-Whitney

Тест для случая двух независимых выборок

Пример

Алкоголики лечатся в больнице и получают лечение. Есть обычное лечение (control) и продвинутое (sst). Надо понять, есть ли разница. Тут в скобочках указаны ранги

2 independent samples

Table 4.2 Alcohol Intake for 1 Year (Centiliter of Pure Alcohol)

Control		SST	
1042	(13)	874	(9)
1617	(23)	389	(2)
1180	(18)	612	(4)
973	(12)	798	(7)
1552	(22)	1152	(17)
1251	(19)	893	(10)
1151	(16)	541	(3)
1511	(21)	741	(6)
728	(5)	1064	(14)
1079	(15)	862	(8)
951	(11)	213	(1)
1319	(20)		

Source: L. Eriksen, S. Björnstad, and K. G. Götestam (1986).

Идея: смешать все вместе и посмотреть на ранги.

Формализация

Итак, мы имеет две независимые в совокупности выборки

$$
\begin{gathered}
X_1, \dots, X_n \sim F \\
Y_1, \dots, Y_n \sim G \\
n \neq m
\end{gathered}
$$

Предположение:

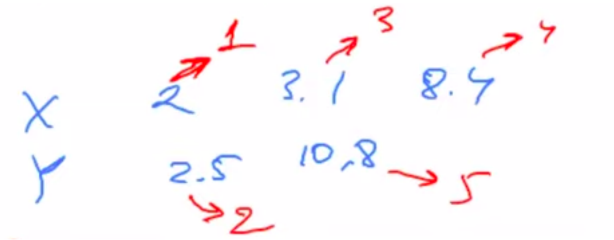
$$F(x) = G(x - \Delta) \;\Leftrightarrow\; x + \Delta \underset{d}{=} Y$$

Это сильное предположение. Но оно конечно слабее любого параметрического предположения.

Гипотеза:

$$
\begin{aligned}
\mathcal{H}_0 &: \quad \Delta = 0 \\
\mathcal{H}_1 &: \quad \text{else}
\end{aligned}
$$

Пример:



Соответственно, как работает тест. Пусть ранги $X_1, \dots X_n$ это $R_1, \dots R_n$, а ранги $Y_1, \dots Y_n$ это $S_1, \dots S_n$. Тогда нужно посчитать такую статистику:

$$W = \sum_{j=1}^{m} S_j - \text{Wilcoxon statistics}$$

Название странно, но как есть. Однако существует статистика Манни-Уитни:

$$U = \sum_{j=1}^{m} \sum_{i=1}^{n} \mathbb{I}[X_i < Y_j] - \text{Mann-Whitney statistics}$$

Они связаны тождеством:

$$W = U + \frac{m(m+1)}{2}$$

Асимптотическая теорема

$$\Delta = 0 \Rightarrow \frac{W - \frac{m(n+m-1)}{2}}{\sqrt{\frac{nm(n+m+1)}{12}}} \xrightarrow[n,m\to\infty]{d} N(0,1)$$

Kruskall-Wallis

Тест для нескольких независимых выборок

Пример

Есть медицинская процедур для определения болезни органов дыхания.

Table 6.1 Half-Time of Mucociliary Clearance (h)

	Subjects with	
Normal subjects	Obstructive airways disease	Asbestosis
2.9 (8)	3.8 (13)	2.8 (7)
3.0 (9)	2.7 (6)	3.4 (11)
2.5 (4)	4.0 (14)	3.7 (12)
2.6 (5)	2.4 (3)	2.2 (2)
3.2 (10)		2.0 (1)
$R_1 = 36$	$R_2 = 36$	$R_3 = 33$

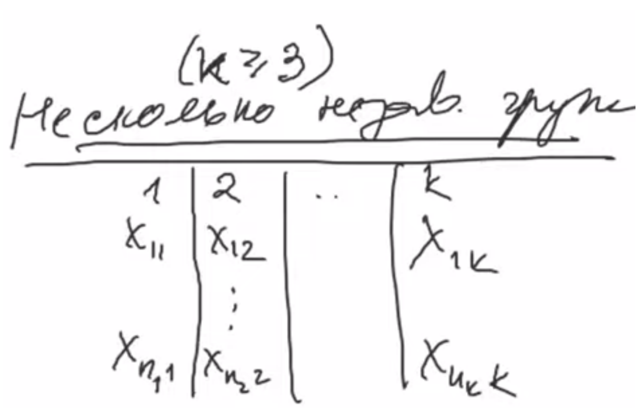
Source: M. L. Thomson and M. D. Short (1969).

Есть три группы. Здоровые, одна болезнь и вторая болезнь. Надо сказать, верно ли, что измеряемый показатель одинаковый во всех группах (или наоборот разный).

Другой пример: проверка эффективности лекарства. Опять три группы: плацебо, старое лекарство, новое лекарство. Надо доказать, что новое лекарство лучше остальных двух групп.

Формализация

Несколько ($\geqslant 3$) независимых групп.



Модель выглядит так:

$$x_{ij} = \Delta + \Delta_j + \varepsilon_{ij}$$

где

- Δ - общая медиана
- Δ_j - медиана по группе
- ε_{ij} - шум

Гипотеза:

$$\begin{aligned} \mathcal{H}_0: &\quad \Delta_1 = \dots = \Delta_k \\ \mathcal{H}_1: &\quad \text{else} \end{aligned}$$

Подсчет статистики. Пусть R_{ij} - ранг в общей выборке. Посчитаем средний ранг по группе (по столбцу):

$$R_j = \frac{1}{n_j} \sum_i R_{ij}$$

Асимптотическая теорема

Теорема. Если выполнена $\mathcal{H}_0$, то

$$\frac{12}{N(N+1)} \sum_{j=1}^{k} n_j (R_j - \frac{N+1}{2})^2 \to \mathcal{X}^2_{R-1}$$

По сути этот тест является обобщением теста Манни-Уитни

Вопрос

У нас большинство тестов (например, Вилконсона) умеют доказывать что выборки разные: если p-value больше α, то мы отвергаем гипотезу; иначе же ничего не делаем. А что делать в противоположной ситуации - если мы хотим доказать, что выборки одинаковые? Как я понимаю, нужны другие тесты